MULTIPLE STATES IN MACROMOLECULES
I. Qualitative model for a single nucleation process

A. DANCHIN

Institut de Biologie Physico-chimique, rue P. Curie, Paris Ve, France

Received 18 October 1971

1. Introduction

Biological macromolecules assume a most important fraction of their functional structure thanks to chemical interactions much weaker than the covalent bond; among these there are the hydrogen bond and the stacking interaction (due to Π - Π or dipolar interaction, and Van der Waals interactions). The Gibbs free energy involved in such interactions is often not much higher than the 3/2 kT energy of the Brownian motion of the solvent. Thus each association can be broken and remade easily, allowing a breathing of the molecule. Actually, when an association is broken it generally has several possible conformations and the likeliness of reassociation is mainly dependent on the neighbouring associations so that the breathing of the macromolecular structure is a cooperative phenomenon. To describe the multiple possible states of a macromolecule population, Zimm has developed a formalism which is used in most general cases [1–3]. This theory describes the transitions which occur in a polypeptide or a nucleic acid double helix by statistical mechanics. Any interaction can be the beginning of a cooperative transition (a nucleation center) and the cooperativity is reflected by a correlation factor which measures the probability of not forming an interaction when the neighbouring one is present. This factor is considered as temperature-dependent for polypeptides. A macromolecular population is thus in an equilibrium between multiple states which depends on the energy for each interaction and on the correlation factor value.

We shall try here to describe a somewhat different aspect of nucleation: what happens in the case where only one center of nucleation begins the structural transition and when the correlation factor is strongly temperature-dependent? Such a phenomenon is likely to occur in molecules when an important center must assume a well-fixed structure over a large range of temperatures whereas the external part of the structure may be more labile. A rigid kernel would thus be built in one nucleation process and its temperature-dependent rigidity would extend farther at low T than at high T. Such a macromolecule should thus be in equilibrium between multiple states, the states differing by the internal fine structure of the kernel. Such a model is therefore relevant to the case of proteins, which have the so-called "globular" structure, and to many tertiary structures of macromolecules; it will be illustrated experimentally in a second article on the special case of transfer RNA [4].

2. The theoretical model

The macromolecule, M, is represented by a finite set of kernels ordered by inclusion (one can eventually go to infinity by continuity). Each kernel k_i can either be in a rigid or non-rigid state and the outer rigid kernel envelopes only a set of rigid kernels k_j, $j > i$. M will be described as a progressively melting set of kernels. At low T all the kernels are rigid and the outer one is k_1; at high T all are non-rigid and the last one is k_n (corresponding to the temperature T_n).

We shall now add a new aspect to each kernel: we

assume that it can be in two possible states: open (0) or closed (1). The i^{th} kernel will be noted k_1^0 or k_i^1. This notation corresponds to the presence or absence of a specific interaction: either primary (for instance, 2′ or 3′ endo structure of a ribose residue), secondary (for instance, formation of a specific hydrogen bond, leading to a well-defined helix) or tertiary (local folding due to stacking, other polar interactions, or to binding an ion in a specific place). The external aspect is described by the state of the set of all the kernels included in k_i : we shall write it $K_i^{(\delta j)}$ where δ_j is either 0 or 1 according to the state of the j^{th} kernel $i < j < n$.

Let us imagine now that we use an experimental way, E, for removing the molecules which have their external rigid kernel open. We thus put the system out of equilibrium and we can try to describe what happens in such case when the temperature is changed.

We consider the case where E is fast enough to allow us to neglect the multiple re-equilibrium process when the population has all its kernels k_i^1 in the closed state, and we separate the problem in two parts: first, the $k_i^{\delta} i$ distribution is obtained as a function of temperature by the direct use of E, and second, we express the re-equilibrium between multiple states after E has been used. This is reflected by the $K_i^{(\delta j)}$ general interconversion after the equilibrium is perturbed (as after the use of E when $\delta_i = 1$)

Let us ascribe an energy level to each i value and split this level into two sublevels corresponding to the open and closed kernel, with a Gibbs free energy difference Δg_i. We shall say that the level i is accessible at the temperature T when this level corresponds to the outer rigid kernel.

As a formal representation of the population of M we shall suppose that the number of open kernels follows a law of probability $\pi(i)$ increasing as i increases; the law of accessibility is $i(T)$ therefore $\pi(i) = \pi(i(T)) = P(T)$. This is consistent with the phenomenological point of view that the number of open kernels increases among those accessible at higher temperature (between the values 0 and 1); we shall see later a form for this law. This hypothesis implies that the levels accessible at the higher temperatures are first distributed when M assumes its structure in solution. This is understandable since these levels correspond to the most stable part of the tertiary structure, which is destabilized only at the highest temperatures.

If i and $i + k$ are two levels (accessible at two different temperatures), we shall suppose that when k is small the open state of i is completely dependent on the state of $i + k$ and that, when k is much bigger, the state of the two levels are independent. Thus the open state will be a function of a correlation factor varying between 0 and 1. Expressed in a phenomenological way this means that the structure of the neighbouring fragments, and that this dependence becomes less and less important as the distance between fragments increases.

Let T_1 be the highest temperature at which all the outer kernels are closed (E removes nothing); let us take T_1 as reference. The set of energy levels accessible for $T > T_1$ will be n, and we shall study the distribution of open states among those levels. Before we give a mathematical representation of the probability law $P(T)$ of finding the i^{th} kernel open at the temperature T, we shall discuss the thermal behaviour of the M population in the most general case.

2.1. *Distribution of open kernels in function of T*

This distribution is observed thanks to E: the level i which is accessible at the temperature T corresponds to an open kernel with the probability:

$P(T) = \pi(i(T))$ with $0 < P(T) < 1$ increasing with T, and

$P(T) = 0$ for $T < T_1$ and $P(T) = 1$ for $T < T_n$.

Thus if $T < T_1$ E removes no molecule from the mixture. If $T_1 < T < T_n$ E removes a fraction $P(T)$ of the macromolecular population due to the fact that such a fraction has its outer kernel open. If now $T > T_n$ E removes all the molecules since they all have their external kernel open.

In the case of a stepwise use of E we shall obtain the following results:

a. Measure of the effect of E at T_1 followed by a measure at T_{i+k} $(0 < i < n; k > 0)$: at T_i the level accessible is i and E removes $P(T_i)$ molecules; at T_{i+k} the new level is $i + k$ so that the final value of the measurement by E is $P(T_{i+k})$: the new proportion between the two temperatures is therefore:

$$P_1(T_{i+k}) = \frac{P(T_{i+k}) - P(T_i)}{1 - P(T_i)} \quad .$$

This is true only in the case where one can neglect the re-equilibrium between multiple states during the time used for E.

b. The same experiment with $k < 0$: the phenomenon is more involved since the M population has been separated at T_i higher than T_{i+k} and the distribution of open states might change during the passage from i to $i + k$ (due to new distribution of hydrogen bonds, exchange of ions, or other conformational changes). We shall therefore suppose that the new distribution will depend on the former one with a correlation factor: $\rho = 0$ means that the new distribution is independent of the previous one and the M population behaves exactly as if it had not been previously submitted to E; on the other hand, when $\rho = 1$, the new distribution of open kernels is completely dependent on the former distribution, so that one finds exactly the state corresponding to the previous use of E, i.e. no change when E is used at this single temperature (E no longer removes molecules from the mixture): this gives an apparently irreversible pattern. In any case the new probability of sorting out molecules is:

$$P_1(T_{i+k}) = \mathrm{Sup}\,(0, \frac{P(T_{i+k}) - \rho\, P(T_i)}{1 - \rho\, P(T_i)})$$

$$k < 0 \qquad 0 < \rho < 1.$$

2.2. *Determination of P(T)*

The Gibbs free energy difference between the open and closed state of the level i is Δg_i so that the proportion of molecules in the open state can be evaluated by the classical distribution:

$$P(T) = \frac{\text{molecule in state (0)}}{\text{total number of } M} = \frac{K \exp(-\Delta g_i/RT)}{1 + K \exp(-\Delta g_i/RT)}$$

in the case where the re-equilibrium process between k_i^0 and k_i^1 is slow enough. The real form for $P(T)$ must be obtained for each experimental case: as a first, simple approximation one can integrate the k frequency factor in the entropic fraction of Gibbs free energy and eventually assign the same enthalpy and entropy to each i value.

2.3. *Slow equilibrium process for $K_i^{0(\delta j)}$ to $K_i^{1(\delta j)}$ interconversion*

Since the distribution of open states at the i^{th} level is correlated to the distribution of states of all the levels accessible at the higher temperatures, to say that the i^{th} kernel goes from the closed state to the open state means that the whole set of kernels has to change; this must be observed in the M population by a slow (because multiple) equilibration process. If the correlation between the state of the levels is total, the Gibbs free energy Δg_i involved in the interconversion would be $\Delta g_i = \sum_{i=1} \Delta g_i$; the probability of the interconversion may be thus very low (at least when T is low enough) so that the equilibrium, at least for $i \ll n$, is attained very slowly. The measurement of the re-equilibrium process after E has been used can give a method of evaluating the number n of the different levels and also of the correlation between them.

3. Discussion

We have shown a qualitative approach for analyzing a single nucleation process in macromolecules and it will be illustrated in a next article on the special case of transfer RNA; we shall emphasize here the differences between this model and the classical Ising–Zimm models [1–3, 5].

The first obvious differences appear in the starting hypotheses since the I–Z model assumes a random and multiple nucleation process; moreover, in this model the correlated interactions attached to one nucleation center are independent of the temperature (this is true in the polypeptide model; in the nucleic acid model the correlation factor is assumed to be temperature-dependent because the equilibrium constant for a stacked base pair is strongly temperature-dependent [6]. We have, on the contrary, supposed one nucleation center and a pattern for the rigidity around this center depending on the temperature. Thus, in the I–Z model an equilibrium must appear either to be reversible (i.e. the use of E proceeds to remove all the molecules, since the equilibrium is completely displaced by E) or, at least, frozen (i.e. after E has removed part of the molecules

and reached a plateau, a single passage to a higher temperature before use of E at the lower T gives a raise to a new separation of M by E, thus showing that the passage to the higher T again has allowed a re-equilibration process). In the model presented above, on the opposite, the second aspect may be completely absent (i.e. the passage to a higher T before using E at the former temperature gives no new separation of the macromolecular population) and this appears as an apparently irreversible phenomenon.

One would expect such a dramatic effect in the case where a nucleation center dominates strongly the folding of the macromolecule, especially in the case where certain topological properties such as knots are found [7].

A more quantitative description of our model may be obtained if one uses the partition function described by Zimm or Lifson albeit in changing the nucleation pattern to one center and giving a temperature dependence to the correlation factor. This would be useful in the case where a population of macromolecules can be described by an interaction with a proper (i.e. sensitive) means of investigation E. We think, however, that at the present time we do not have sufficiently precise experimental data to refine the model, but the gross features of a population may be obtained with simple assumptions; for instance, concerning the Gibbs free energy differences between the open and closed states of the kernels (e.g. Δh_i and Δs_i independent of i).

The last important feature in the case of a single nucleation process, such as the one we described, is that a macromolecular population should be in an equilibrium between multiple states with a fast component due to the external structure of the rigid kernel, and a slow component due to the internal structure of the kernel surrounding the nucleation center.

Acknowledgement

I am highly indebted to Dr. R.L. Baldwin for encouragements and critical comments on the manuscript.

References

[1] B.H. Zimm and J.K. Bragg, J. Chem. Phys. 28 (1958) 1246; 31 (1959) 526.
[2] B.H. Zimm, J. Chem. Phys. 33 (1960) 1349.
[3] Review by H. Scheraga, in: The Proteins, Vol. 1, ed. H. Neurath (Acad. Press, New York, 1963).
[4] A. Danchin and M.N. Thang, FEBS Letters 19 (1971) 297.
[5] E. Ising, Z. Physik 31 (1925) 253.
[6] R.L. Baldwin, Review for Accounts of Chemical Research, in press.
[7] J. Ninio, Biochimie 53 (1971) 485.
[8] S. Lifson, J. Chem. Phys. 40 (1964) 3705.